

药物真实世界研究设计与方案框架 指导原则（征求意见稿）

2022年6月

目 录

一、概述.....	1
二、真实世界研究设计的主要类型.....	1
(一) 观察性研究设计.....	2
(二) 实用临床试验设计.....	4
(三) 单臂研究设计.....	6
三、真实世界研究方案的主体框架.....	8
(一) 方案摘要.....	9
(二) 研究背景.....	9
(三) 研究目的.....	9
(四) 研究假设.....	10
(五) 整体设计.....	10
(六) 研究人群.....	10
(七) 治疗或干预.....	11
(八) 研究终点.....	12
(九) 基线变量及重要协变量.....	13
(十) 观察期/随访期与观测/随访时间点.....	14
(十一) 数据治理/数据管理计划.....	14
(十二) 偏倚考虑.....	14
(十三) 统计分析计划.....	15
(十四) 质量控制.....	20
(十五) 伦理.....	21
(十六) 注册登记.....	21
(十七) 方案修订.....	21
(十八) 组织实施.....	21
四、真实世界研究设计的其它考虑.....	21
(一) 真实世界研究路径的可行性.....	21
(二) 目标人群的代表性.....	22
(三) 混合型研究设计.....	22
(四) 虚拟对照方法.....	23

(五) 估计目标.....	24
(六) 模仿目标临床试验	26
五、与监管机构的沟通.....	26
参考文献.....	28
附录 1 名词解释.....	30
附录 2 中英文词汇对照.....	34

1 药物真实世界研究设计与方案框架指导原则

2 (征求意见稿)

3 一、概述

4 已发布的《真实世界证据支持药物研发与审评的指导
5 原则（试行）》和《用于产生真实世界证据的真实世界数
6 据指导原则（试行）》为真实世界证据支持药物研发与审
7 评奠定了基础。采用真实世界研究支持儿科药物、罕见疾
8 病药物等监管决策的指导原则也相继发布。

9 为了指导申办者科学合理的设计真实世界研究，明确
10 真实世界研究方案撰写的技术要求，本指导原则将重点阐
11 述药物研发中真实世界研究设计以及研究方案制订的基本
12 考虑，为药物研发中开展真实世界研究提供指导意见。

13 本指导原则适用于通过真实世界研究获得药物评价的
14 临床证据。真实世界证据支持药物研发和监管决策的适用
15 情形参见《真实世界证据支持药物研发与审评的指导原则
16 （试行）》。本指导原则也可供以非注册研究为目的的真
17 实世界研究参考。

18 二、真实世界研究设计的主要类型

19 真实世界研究设计包括观察性（或非干预性）研究设
20 计和干预性研究设计（如实用临床试验）。单臂研究设计

21 是一种特殊的设计形式，其研究组可以是干预性的，也可
22 以是观察性的，其外部对照通常基于真实世界数据而设定。

23 (一) 观察性研究设计

24 观察性研究可分为队列研究、病例对照研究和横断面
25 研究。以因果推断为目的的观察性研究建议采用队列研究
26 设计。本指导原则后文若无特别说明，所述观察性研究均
27 是指队列研究。

28 根据研究方案中定义的研究起始时间和研究数据的产
29 生时间，队列研究可分为回顾性、前瞻性和回顾前瞻性队
30 列研究。回顾性队列研究收集的是历史数据，即研究开始
31 前生成的数据；前瞻性队列研究收集的是研究开始后的数
32 据；回顾前瞻性队列研究既收集已有的历史数据，也收集
33 研究开始后的数据。

34 队列研究设计主要考虑目标人群队列、因果推断分析
35 和质量控制三个方面。其它方面的考虑在下文的研究方案
36 部分有具体阐述。

37 1. 目标人群队列

38 目标人群队列根据临床所关心的问题而定，具体以数
39 据体现，即目标人群从研究的治疗开始到观察期结束所形
40 成的纵向观测数据。目标人群队列的具体定义应基于研究
41 目的、入排标准（包括初治者（new user，本指导原则在实

42 操上定义为：纳入研究队列之前在充分的洗脱期内未使用
43 研究用治疗的病例）或非初治者（non-new-user）、数据来
44 源和数据治理/管理计划综合考虑。数据来源通常是多中心
45 的，对于单中心情况，需充分评估该中心人群的代表性及
46 研究结论的外推性。目标人群所收集的重要变量包括治疗
47 （包括研究队列和对照队列）、基线、协变量、时依变量和
48 结局变量等。观察性研究的样本量应在充分考虑混杂因素、
49 缺失数据等因素的基础上满足统计假设的要求，通常不设
50 上限，特别是回顾性研究。队列起始时间、观察期/随访期
51 的长短和观测时间点/访视点的确定应符合所研究疾病的特
52 征和临床评价要求。

53 2. 因果推断

54 观察性研究由于变量间因果关系的不确定性和复杂性
55 使得因果推断具有挑战。不同分析模型的选择往往会导致
56 分析结果不同，因此，为了避免结果驱动的偏倚，需要在
57 设计阶段明确主分析将要采用的分析数据集、分析模型及
58 其相对应的统计假设。为使研究结果更为准确和稳健，应
59 考虑混杂偏倚、选择偏倚、信息偏倚等重要偏倚识别及控
60 制方法，以及缺失数据的处理策略及其基于的假设；还应
61 针对可能影响研究结果的各种因素，如模型假设背离或各
62 类潜在偏倚来源，充分考虑敏感性分析及定量偏倚分析计

63 划和策略。

64 3. 质量控制

65 质量控制的主要目的是保证获得高质量的分析数据。

66 一方面，需要事先制定数据治理计划（针对历史数据）或
67 数据管理计划（针对前瞻收集数据），保证所产生的数据能
68 够满足适用性要求（参见《用于产生真实世界证据的真实
69 世界数据指导原则（试行）》）；另一方面，应制定具体措施
70 保障观测变量值的准确性，例如在保障测量工具、度量单
71 位和评价方法的一致性方面的具体措施。

72 （二）实用临床试验设计

73 实用临床试验(pragmatic clinical trial, PCT)又称实操临
74 床试验或实效临床试验，是指尽可能接近真实世界临床实
75 践的临床试验，是介于随机对照临床试验（RCT）和观察性
76 研究之间的一种研究类型，属于干预性研究。与 RCT 不同
77 的是：PCT的干预既可以是标准的，也可以是非标准的；既
78 可以采用随机分组方式，也可以自然选择入组；受试病例
79 的入选标准可以相对较宽泛；对干预结局的评价不局限于
80 临床有效性和安全性；PCT更多地使用临床终点，而很少使
81 用传统 RCT 中可能使用的替代终点；可以同时考虑多个治
82 疗组，以反映临床实践中不同的标准治疗，或设置多个剂
83 量组达到剂量探索目的；一般不设安慰剂对照；如果困难

84 以实施而不采用盲法，应考虑如何估计和控制由此产生的
85 测量偏倚；数据的收集通常依赖于患者日常诊疗记录，但
86 也可以设置固定的随访时间点，其时间窗通常较 RCT 更
87 宽。与观察性研究不同的是，PCT是干预性研究，尽管其干
88 预的设计具有相当的灵活性。

89 PCT设计应重点考虑以下因素：①收集到的数据是否适
90 用于支持产生真实世界证据；②治疗领域和干预措施等是
91 否符合各种形式的常规临床实践；③是否具有足够的可以
92 用于评价的病例数（特别是临床结局罕见的情况）；④参
93 与 PCT 的各试验中心甚至不同的数据库之间对终点的评价
94 和报告方法是否一致；⑤是否采用随机化方法控制偏倚；
95 ⑥当盲法不可行时，应考虑非盲对结局变量，特别是患者
96 报告的结局（patient reported outcome, PRO），可能产生的
97 影响，可使用不受治疗分组影响的客观终点（如中风、死
98 亡等），以减少非盲可能带来的偏倚；⑦分析方法的考虑
99 可参照观察性研究的分析方法。

100 对于随机化的 PCT（P-RCT），还需要特别阐明治疗策
101 略的选择（如单次治疗策略或持续治疗策略）和有效性的
102 主分析所基于的数据集。由于 P-RCT 在随机化之后出现的
103 治疗策略更改、剂量改变、停药、转组、数据缺失等情况
104 较 RCT 更为普遍，因此，相较于 RCT 通常基于 ITT/mITT

105 （调整 ITT）进行主分析，P-RCT 则需要考虑基于符合方案
106 数据集（Per-protocol Set, PPS）是否更为合理的问题，或
107 者考虑更加合适的数据集定义，并在样本量计算时予以充
108 分考虑。

109 （三）单臂研究设计

110 采用单臂研究首先要考虑的问题是其前提条件是否充
111 分，例如，采用RCT难以实施或具有重大伦理风险，属于危
112 及生命、复发难治、无药可治或甚为罕见的疾病。单臂研
113 究组如果是干预性的，为单臂试验；如果是非干预性的，
114 为单臂观察性研究。无论是干预或非干预的，单臂研究设
115 计通常应设置外部对照，外部对照采用的形式有基于疾病
116 自然史队列数据或其他外部数据的历史对照或平行对照，
117 或者目标值对照。为了减少偏倚，采用外部对照需考虑其
118 目标人群特征（人口学、基线水平和临床特征等）、诊断
119 和治疗标准、伴随治疗、结局的测量和评价标准等对结局
120 （预后）有潜在影响的各种因素与研究组是否足够相似，
121 以保证与研究组有较好的可比性。此外，单臂研究设计至
122 少还应考虑以下内容。

123 1. 研究组设置

124 研究组的设置主要分干预性和非干预性，前者更为常用。
125 对于干预性设计，研究组需要定义标准干预，且在研究实

126 施过程中严格执行所规定的干预措施；对于非干预性设计，
127 研究的治疗通常没有统一标准，且在治疗过程中患者可能
128 会同时接受其它治疗，使得治疗模式较为复杂多样，对此
129 可通过设置合理的入选和排除标准来定义较明确的目标治
130 疗。

131 2. 对照设置

132 (1) 历史对照

133 以既往获得的疾病自然史队列或其它外部真实世界数
134 据作为对照，应考虑人群异质性及不同历史时期对疾病的
135 定义、诊断、分类、自然史和可用的治疗手段等对疗效可
136 比性和一致性的影响。

137 (2) 平行外部对照

138 前瞻性地收集与研究组同期的疾病自然史队列或其它
139 外部真实世界数据作为对照。

140 (3) 目标值对照

141 目标值的确定应有充分依据，优先依次考虑国家标准、
142 行业标准和专家共识，否则，需要根据已有的相关信息，
143 包括但不限于公开发表的文献、研究报告、相关研究的原
144 始数据等，通过综合分析确定目标值。

145 (4) 混合对照

146 将既往及研究同期获得的外部数据混合在一起形成对
147 照臂。这些外部数据可以是日常的病例记录，也可以是过

148 去开展不同的临床研究（观察性或干预性的）所获得的数
149 据。研究开始前需评估外部数据的适用性、代表性和预先
150 设定不同部分数据合成时的权重系数，建议预先设置敏感
151 性分析评估混杂因素、不同权重系数等对研究结论的影响。

152 3.其它考虑

153 采用外部对照的单臂研究由于混杂因素、人群异质性
154 和各种可能偏倚的影响，因果推断结论具有较大的不确定
155 性。为克服或减少这些局限，除上述考虑外，还应注意：

156 ①主要终点采用客观指标，如肿瘤临床研究的客观缓解率
157 （ORR）；②明确并严格把握入组人群的入排标准及筛选
158 过程；③要确保所采集的数据符合真实世界数据的适用性
159 要求；④较之于历史对照，更鼓励采用平行外部对照；⑤
160 事先恰当地定义主分析的统计分析方法，如合理利用多因
161 素模型、倾向评分（propensity scores, PS）方法，虚拟对
162 照（virtual control）方法、工具变量方法等；⑥若对照组选
163 择或主分析模型采用基于匹配的方法，应在方案中事先明
164 确匹配标准；⑦要充分使用敏感性分析和偏倚的定量分析
165 来考察未知或未测量的混杂因素、效应异质性、模型假设
166 不成立以及其它各种可能偏倚对分析结果的影响。

167 三、真实世界研究方案的主体框架

168 不同设计类型的真实世界研究方案的主体框架基本相

169 同，个别不同之处将在主体框架下相应内容中分别阐述。

170 以下是真实世界研究方案的建议框架，但并不排除个别研
171 究项目的某些特殊考虑。

172 （一）方案摘要

173 以表格形式摘录研究方案的主要内容，突出重点，力
174 求简洁。主要内容包括：标题、研究目的、研究假设、整
175 体设计、研究人群（含诊断标准、入选标准、排除标准、
176 剔除标准等）、治疗（定义研究组和对照组）、研究终点、
177 基线变量与重要协变量、安全性指标、观察期与观测时间
178 点、数据来源、数据治理或数据管理、样本量及其确定依
179 据、统计分析、偏倚控制等。

180 （二）研究背景

181 简要介绍研究背景，包括国内外研究的现状和意义，
182 本研究的前期基础等。应充分论述选择真实世界研究路径
183 的主要依据（如其它研究类型的不可行性和伦理风险等），
184 以及本研究的定位，例如用于注册的关键支持证据、辅助
185 支持证据、或基于探索目的的研究等。

186 （三）研究目的

187 根据目标人群、治疗（含对照）和结局，简要阐述研
188 究目的，即本研究计划回答的临床科学问题，明确主要目
189 的和次要目的（如果有），也可包括探索性目的。

190 (四) 研究假设

191 根据研究目的提出研究假设。

192 (五) 整体设计

193 简述研究的整体设计，包括多中心或单中心、观察性
194 或干预性、单臂或双臂/多臂等要素。若是观察性研究，应
195 说明是回顾性的还是前瞻性的。

196 干预性研究应说明是否采用随机化分配，如果采用，
197 应详细说明具体的随机化分配方法及实施过程；是否采用
198 盲法，如果采用盲法（单盲或双盲），应说明具体实施办法；
199 如果采用开放设计，需说明终点事件是否采用盲评，若采
200 用盲评，如何实施。

201 单臂研究应说明研究组是干预性还是观察性的，以及
202 采用何种形式的外部对照。

203 (六) 研究人群

204 1. 诊断标准

205 如果所研究的疾病有不同诊断标准，应该说明本研究
206 所采用的诊断标准及其出处，并给出所采用诊断标准的具体
207 内容，若内容较多可以在附件中呈现。还应标明疾病代
208 码（如ICD9/ICD10等）。如果诊断标准唯一，且为医学专业
209 人员所熟知，不做说明不会影响研究者对疾病的判断，可
210 以不在方案中提及。

211 2. 入选/排除/剔除标准

212 入排标准及剔除标准的制定应能代表研究的目标人群。
213 一般而言，观察性研究的入排标准较干预性研究宽松。应
214 注意剔除标准可能导致的恒定时间偏倚（immortal-time bias）
215 或选择偏倚，必要时对重要的排除标准做出解释，并评估
216 其对分析结果的影响。

217 （七）治疗或干预

218 本指导原则中，关于采用研究药物或治疗策略的人群
219 队列，对于观察性研究，称之为“治疗组”或“治疗队列”；对
220 于干预性研究（如 PCT），称之为“试验组”。关于采用非研
221 究药物或治疗策略的人群队列，称之为“对照组”或“对照队
222 列”。

223 1. 治疗组/试验组

224 对于治疗组的定义，应阐明具体治疗方法，如药物治
225 疗的剂量、频次、给药途径、疗程等，以及药物的商品名
226 和生产厂家；如果是物理治疗（如放疗或激光治疗），应给
227 出具体的治疗参数。观察性研究中治疗策略和治疗模式由
228 临床实践所决定，因此具有多样性，在数据收集、因果推
229 断和结果解释时应予以考虑。

230 对于试验组的定义，与观察性研究不同的是，治疗方
231 法通常应固定下来，形成相对标准的治疗策略。

232 2. 对照组

233 真实世界研究通常选择阳性对照或标准治疗对照，阳
234 性对照应是目前或数据采集的起止期（如回顾性研究）临
235 床实践中公认的疗效明确的治疗方法或治疗策略。对照组
236 应像治疗组或试验组一样描述具体的治疗方法或策略。

237 除单臂研究的历史对照外，观察性研究对照的选择应
238 与治疗组同时期。对于回顾性研究，为了避免病例选择偏
239 倚，原则上应选择研究所定义的数据采集的起止期内所有
240 的治疗组和对照组的病例，或者采用严格的随机抽样方法
241 选择病例（例如因为已有数据量太大而无法承受治理和分
242 析全部数据的情况）。对于前瞻性研究，对照组的选择标准，
243 特别是与治疗组的匹配方法，应明确定义。

244 干预性研究对照组的选择与 RCT 类似。

245 单臂研究对照的选择见上述第三章的“（三）单臂研究
246 设计”中的“对照设置”一节。

247 3. 伴随治疗

248 真实世界研究中，伴随治疗的情况较为常见，应在方
249 案中尽可能地阐述清楚可能出现的伴随治疗，对于未能预
250 见的伴随治疗，也需要在分析过程中予以充分考虑。

251 （八）研究终点

252 1. 有效性终点

253 应定义主要终点和次要终点，如必要还需定义关键次

254 要终点。有效性终点应完整定义，包括终点的名称、观测
255 的时间点或时间段、测量方法与工具、计算方法、评价方
256 法等。必要时，可设置独立第三方终点事件判定委员会，
257 并描述实施办法，例如所执行的标准操作程序（SOP）。应
258 注意，真实世界研究的主要终点通常不采用替代终点，如
259 采用需充分说明理由。

260 2. 安全性终点

261 安全性终点的考虑可参照常规临床试验。

262 3. 探索性终点（如果有）

263 研究如有需要，也可以设置某个/些探索性终点，例如
264 药物经济学终点等。

265 （九）基线变量及重要协变量

266 研究方案中应明确基线变量和重要的协变量，以及它
267 们的度量单位和观测时间。这些变量的确定依据主要来自
268 对目标研究人群现有的研究成果，例如指南、专家共识、
269 公开发表的文献、会议报告等提及的影响疗效的变量/因素，
270 也有来自项目组专家的认识。重要协变量的确定应具备合
271 理性，可结合各因素间的因果路径图确定，并综合考虑前
272 期数据评估结果。在已确定的重要协变量中，建议在方案
273 中明确协变量的属性，如效应修正因素、危险因素、混杂
274 因素（包括依时混杂因素）、中间变量、碰撞变量、工具变

275 量等。

276 (十) 观察期/随访期与观测/随访时间点

277 应明确研究的观察期或随访期，以及观测对象的观测
278 或随访的起始时间、时间间隔和时间点，合理定义窗口期。

279 (十一) 数据治理/数据管理计划

280 在真实世界研究中，应准确理解数据治理和数据管理
281 的概念。对于既往数据，无论是病历记录的原始数据，还
282 是开展不同临床研究所获得的数据，都应经过统一的数据
283 治理使其满足分析的要求。对于前瞻性收集的数据，应通
284 过严格和规范的数据管理，为研究提供高质量的用于分析
285 的数据。数据治理或数据管理计划如果篇幅较大，可以呈
286 现在附录中，但必须与研究方案同步完成。

287 应明确研究数据来源，包括所来自的研究中心、收集
288 数据的起止时间、数据存储的系统和记录形式。如果是来
289 源于既往的研究，应描述原始数据的记录和存储形式，以
290 保证研究数据可追溯。

291 数据治理和数据安全的具体要求可参见《用于产生真
292 实世界证据的真实世界数据指导原则（试行）》。

293 (十二) 偏倚考虑

294 偏倚是真实世界研究特别需要考虑的问题，在方案中
295 应充分考虑各种潜在偏倚及其影响，并制定控制偏倚的有

296 效措施。常见的偏倚包括：因测量、数据收集或评价方法
297 的不准确或不一致导致的信息偏倚，因选择性地入选和/或
298 排除数据或失访、退出、剔除、记录缺失等导致的选择偏
299 倚，因人群变化、治疗变化、研究背景变化等原因导致的
300 疗效异质性，因分析中未能充分控制混杂因素导致的混杂
301 偏倚，因未事先确定主分析方法而选择采用不同分析方法
302 中最有利的结果导致的结果驱动等偏倚。此外，不同的研
303 究中还可能发生其它具体的信息偏倚，例如，在记录生存
304 时间时可能产生的恒定时间偏倚或领先时间/起点时间偏倚
305 （lead time bias /zero-time shift bias），基于文献的 meta 分析
306 可能存在的发表偏倚（publication bias），回顾性研究中回忆
307 以往事件可能产生的回忆偏倚（recall bias），因入选非初治
308 病例而产生的幸存者偏倚（survivor bias）等。

309 （十三）统计分析计划

310 为了避免结果驱动偏倚和保证研究过程的透明性，真
311 实世界研究特别强调，至少主分析计划应该与研究方案同
312 步确定，这与 RCT 中规定统计分析计划可以在数据库锁定
313 之前完成有很大不同。如果主分析计划篇幅较大，可以附
314 件形式呈现。独立的主分析计划除了摘录方案中的一些关
315 键要素外，如研究目的、目标人群、终点指标及其定义等，
316 在统计分析方面至少应该包含以下内容。

317 1. 样本量估计

318 临床试验的样本量估计通常要考虑的因素有：研究类
319 型、比较类型（优效性或非劣效性）、统计分析方法、结局
320 变量预期的效应量或参数、统计分布、检验水准、单双侧
321 检验、检验效能、分配比例、多重性、脱落剔除率、依从
322 性等。真实世界研究应采用主分析所对应的样本量估计方
323 法，在估计时，除了需要考虑上述因素外，还需考虑混杂
324 因素等的调整对样本量的影响。其常用的方法有两种：①
325 基于多因素模型的估计方法。例如，若主分析采用 logistic
326 回归模型，样本量估计需要的额外参数为协变量与治疗分
327 组的广义决定系数 R^2 。有关多变量的样本量估计方法可参
328 阅相关的专业文献。② 基于匹配的估计方法。例如，若主
329 分析方法采用匹配方法对各因素进行控制，如倾向性评分
330 匹配法，则样本量估计时一般先估计匹配后所需的样本量，
331 然后按照预期的匹配折损率回推最终所需的样本量。

332 此外，还有某些经验方法，例如，基于每个协变量所
333 需阳性事件数（events per variable, EPV）的经验估计方法，
334 根据多因素模型可能纳入的协变量个数的若干倍数估计样
335 本量等。

336 需注意，对于采用外部对照组的单臂试验研究，对照
337 组的样本量通常应不少于治疗组的样本量，或可以数倍于

338 治疗组。另外，观察性研究（特别是回顾性研究）的数据
339 缺失率较高，设计时应留有充分余地。

340 2. 数据集定义

341 真实世界研究的数据来源及其质量有很大差异，而且
342 不同分析回答的问题各异，应根据不同的分析定义不同的
343 数据集，如有效性数据集和安全性数据集。如果涉及随机
344 分配，应基于随机分组定义数据集。如果分析的目标人群
345 是数据集的一个子集，应将子集标记为对应的目标人群。

346 3. 缺失数据处理

347 真实世界研究中，数据缺失较为普遍，甚至缺失比例
348 较大。在数据治理过程中，应尽可能追踪捕捉遗漏的记录，
349 使数据质量有所改善。尽管如此，主分析或敏感性分析中
350 仍面临缺失数据的处理问题，在主分析计划中应阐述缺失
351 数据的处理方法，并说明其理由。

352 4. 描述性分析

353 描述性分析能够刻画变量（特别是基线变量）的主要
354 特征。所有指标/变量，包括终点变量，均应进行描述性分
355 析，所采用的描述统计量应根据变量的分布特征合理选择。

356 5. 异质性分析

357 应事先考虑可能的异质性因素，如研究中心、年龄、
358 性别、病情程度等，为亚组分析或分层分析打好基础。同

359 时阐述异质性的评估方法，如采用何种分析模型；以及异
360 质性的判断标准，例如以 0.10 的检验水准判断分组与潜在
361 异质性因素是否存在交互作用，但应注意异质性判断标准
362 的确定应结合研究目的和临床意义综合考虑。

363 6. 主分析

364 主分析围绕主要终点进行统计分析，是研究结论的最
365 主要依据，应给予详尽和严谨的阐述，包括但不限于：统
366 计假设；非调整分析和调整分析所采用的模型及其所基于
367 的假设；拟纳入调整分析的协变量的初步考虑以及明确在
368 分析过程中根据观测数据筛选各类变量的明确规则，包括
369 混杂因素/时依混杂因素、风险因子、中间变量和潜在异质
370 性因素的识别；如果采用倾向性评分匹配方法，应定义匹
371 配比例、匹配方法及其具体参数设置（如卡钳值），以及匹
372 配的均衡性验证方法；对于生存结局的分析需考虑是否存
373 在竞争风险问题。此外应对模型假设进行必要的验证，如
374 非线性关系、非等比例风险等。需要指出，即使 PCT 设计
375 中采用随机分配策略，其主分析对协变量的考虑仍建议与
376 观察性研究相同，因为在研究的实施过程中，PCT（特别是
377 群随机，cluster randomization）对基线均衡性的控制远不如
378 RCT 严格。有关因果推断的具体方法可参阅《真实世界证
379 据支持药物研发与审评的指导原则（试行）》的附录部分和

380 其它相关专业文献。对于样本量极为有限的情况，可能只
381 能进行描述性分析，此时应做相应说明。

382 7. 亚组分析

383 应根据现有的研究结论和认知、可能的异质性因素等
384 明确定义需要进行亚组分析的因素，此外也可考虑主要协
385 变量中与分组变量交互显著的因素进行亚组分析。有关亚
386 组分析的具体方法可参阅《药物临床试验亚组分析指导原
387 则（试行）》。

388 8. 敏感性分析

389 由于真实世界研究中因果推断结论具有不确定性，因
390 此结论的稳健性尤其重要，充分的敏感性分析有助于判断
391 结论的稳健性。敏感性分析应根据不同的假设情景展开，
392 这些情景包括但不限于：缺失数据的不同处理机制；不同
393 的数据集定义；不同的分析方法；模型中不同的协变量组
394 合；未知或不可测的混杂因素影响等。

395 9. 定量偏倚分析

396 偏倚对研究结论的影响是因果推断中需要特别考虑的。
397 鼓励对于各种可能的偏倚，明确判断其是否存在，或设置
398 相关偏倚参数值或分布的方法，并基于偏倚的定量分析考
399 察其对结果的影响，例如，根据剔除标准，分别对剔除病
400 例的数据集和不剔除病例的数据集进行分析，比较其结果

401 的差异以判断是否存在选择偏倚；根据混合型研究中内部
402 数据与外部数据间的疗效差异，判断是否存在效应异质性
403 偏倚，并根据结果设置异质性偏倚参数分布进行校正分析。
404 偏倚参数的分布反映偏倚的大小和不确定性，利用临界点
405 分析考察各种可能偏倚的影响也可视为定量偏倚分析的一
406 种方法。另外，敏感性分析与定量偏倚分析亦可合并描述。

407 10.安全性分析

408 真实世界研究，特别是回顾性研究，对安全性事件的
409 主动监测存在较明显的不足，可能需要提供某些外部证据
410 以弥补其不足，例如研究药物在其它研究中的安全性信息
411 和不良反应监测系统报告的信息。如果研究假设要回答研
412 究药物比对照药物具有更好的安全性，还应提供充分的对
413 照药物的安全性信息。对于主要研究目的是回答安全性问
414 题的情况，可参阅相关指导原则或文献。

415 需要指出，上述与研究方案同步的主分析计划主要是
416 呈现事先确定的将要做的各种分析以及这些分析的假设和
417 条件，而与分析对应的结果的具体输出形式（统计图表）
418 可以在正式的数据分析之前确定。

419 （十四）质量控制

420 一般而言，真实世界研究的质量控制与 RCT 类似外，
421 但需要特别关注数据治理过程的质量控制，具体可参见

422 《用于产生真实世界证据的真实世界数据指导原则（试
423 行）》。

424 （十五）伦理

425 真实世界研究的伦理要求可参照 RCT 的要求，只是在
426 患者知情同意方面，回顾性观察性研究有可能采用泛知情
427 同意形式豁免患者的知情同意。

428 （十六）注册登记

429 应描述本研究在公共网站注册登记的情况。

430 （十七）方案修订

431 在真实世界研究的实施过程中，如果更改数据治理计
432 划或统计分析计划中的主分析计划，属于方案的重要变更，
433 需修订方案，并将修订方案与监管部门充分沟通以达成一
434 致。

435 （十八）组织实施

436 可参照一般的临床研究方案，并根据实施项目的特点，
437 制定实施计划。

438 四、真实世界研究设计的其它考虑

439 （一）真实世界研究路径的可行性

440 在进行研究设计前，应首先对采用真实世界研究路径
441 的可行性进行评估，包括但不限于以下考虑：①传统 RCT
442 是否不可行；②是否有比 RCT 更好的研究路径；③真实世

443 界数据是否足以支持将要开展的研究。无论是数据的质量
444 还是数量（样本量）应均能支持统计分析并产生真实世界
445 证据；④该项真实世界研究在药物研发中的定位，明确该
446 研究所形成的证据在整个证据链中的作用；⑤是否得到监
447 管方面的认同。申办方应就采用真实世界研究路径的可行
448 性方面与监管机构充分沟通，在获得监管机构认可后方可
449 实施。

450 （二）目标人群的代表性

451 观察性研究中目标人群的代表性非常重要，理想的确
452 定研究人群的方法是采用严格的随机抽样。然而，由于临
453 床研究的实际情况，研究人群通常采用的是便利抽样的方
454 法确定的，因此，应充分评估研究人群与目标人群的特征
455 是否存在异质性，及其对研究结论外推性（即外部效度）
456 的潜在影响。

457 （三）混合型研究设计

458 本指导原则中，混合型研究是指同时基于内部的干预
459 性研究数据和外部的真实世界数据，混合形成研究臂和/或
460 对照臂而开展的研究。就此含义而言，采用外部对照的单
461 臂试验设计属于混合型研究设计中的一种类型。混合型研
462 究设计还可用于RCT或PCT。混合型研究设计的关键是将内
463 部数据与外部数据的合并应基于合理的统计模型和方法，

464 以保证内部人群与外部人群特征相一致为原则，将外部数
465 据根据个体水平或整体水平、匹配或赋权等方式与内部数
466 据进行融合，并尽可能进行充分的敏感性分析和定量偏倚
467 分析。若采用基于贝叶斯理论的方法，还应配合充分的针
468 对先验分布和其它相关参数设置的模拟分析。由于内外部
469 人群特征重叠程度及效应一致性程度都会影响外部数据所
470 能够提供的有效样本量，因此，混合型研究估算所需样本
471 量时应确保当前试验纳入足够数量的受试者，使分析结果
472 达到稳健和可靠。

473 (四) 虚拟对照方法

474 虚拟对照方法主要用于无特效治疗条件下单臂试验的
475 有效性评价。其基本思路是反事实论证，即：首先，基于
476 现有的真实世界数据，根据单臂设计所考虑的重要变量，
477 建立不用试验药物的预后预测模型；然后，将单臂试验所
478 获得的协变量数据代入预测模型，计算得到假设不用试验
479 药物的预后预测结果（即虚拟对照）及其汇总统计量；最
480 后计算用试验药物实际观测到的结局的汇总统计量，并与
481 假设不用药的汇总统计量进行比较，从而做出是否有效的
482 推断。该法的关键是建模需要有较大的样本量，以及所建
483 模型要有较高的准确性和稳健性。

484 (五) 估计目标

485 ICH E9 (R1) 将临床试验中构建估计目标归纳为五个重
486 要属性，即目标人群、治疗、终点、伴发事件和汇总统计
487 量。真实世界研究中，如何构建估计目标目前仍处于探索
488 阶段，但与传统的RCT相比，还需考虑一些更复杂的问题，
489 例如，研究人群的异质性、治疗方法的灵活性、伴发事件
490 的多样性、终点选择的特殊性、敏感性分析的复杂性等。
491 现阶段，本指南对真实世界研究中围绕估计目标的主线进
492 行设计并无特定要求，但也鼓励真实世界研究设计中积极
493 探索估计目标理念实施的可行性。以下是真实世界研究设
494 计中估计目标需要特别考虑的问题。

495 1. 研究人群的异质性

496 由于真实世界研究中的入选标准较为宽松，且不采用
497 随机化或不能严格按照随机分配方案实施，故人群异质性
498 是普遍存在的，其来源除了人口学特征、临床特征、地域
499 和研究中心等外，还可能包括不愿意参加 RCT 的患者或
500 RCT中代表性不足的患者群体（例如，少数民族、老年人和
501 居住在偏远地区的人。因此，在疗效估计中应考虑所适用
502 的目标人群。

503 2. 治疗方法的灵活性

504 真实世界研究中，对患者的治疗策略通常有较多的选
505 择，如不同的治疗剂量、合并治疗、疗程等。患者的治疗

506 依从性以及治疗选择的偏好等，都是定义治疗时需要考虑的。
507

508 3. 伴发事件的多样性

509 真实世界研究中，除了需要考虑RCT中常见的因安全性、
510 不耐受性、缺乏疗效等引起的伴发事件和终点伴发事件
511 （如死亡）外，还应考虑因患者行为因素（如患者对治疗的
512 的偏好、治疗的便捷性、医患关系等）和非行为因素（如
513 医保政策的调整影响现在接受的治疗药物、健康状况改善
514 等）导致治疗中断的伴发事件，它们对疗效的影响是不同的。
515

516 4. 研究终点选择的特殊性

517 真实世界研究一般选择临床终点而非替代终点，最好
518 选择单一测量、易观测的临床结果（例如死亡或住院）；
519 复合临床终点的使用应确保构成复合事件的每一组成部分
520 能够获得有效记录，否则要慎重考虑其可行性，例如使用
521 量表工具时，某一维度或条目记录的严重缺失，可能导致
522 整个量表的评价结果失效。

523 5. 敏感性分析的复杂性

524 真实世界研究由于需要控制混杂和偏倚而使得因果推断
525 具有复杂性，同时为使估计量更为准确和可靠（稳健），
526 也对敏感性分析提出了更高的要求。

527 此外，真实世界研究中还有许多面临挑战的问题，如

528 混合研究的数据融合问题，观察性研究中个体生存时间记
529 录的区间删失等，有待深入研究和探索。

530 (六) 模仿目标临床试验

531 模仿目标临床试验 (target trial emulation) 是一种真实
532 世界研究方法，即基于现有的真实世界数据，模仿一个良
533 好的 RCT 设计 (明确的入排标准、治疗策略、治疗分配方
534 法、随访期限和时间点、终点事件评价和统计分析计划等)，
535 产生一个真实世界研究的子集，通过因果推断得出研究结
536 论。该法有助于思考、识别和避免不必要的偏倚，如恒定
537 时间偏倚、非初治者偏倚等，并提供合理的方法来阐明我
538 们在观察性研究中可能需要做出的权衡。模仿目标临床试
539 验要考虑使用的场景，该法的前提是需要拥有非常大样本
540 量的适用真实世界数据，并且具有较高的 RCT 的仿真度，
541 目前在应用方面有待获得共识，但不失为一种值得探索的
542 方法。

543 五、与监管机构的沟通

544 真实世界研究中，研究方案及其实施过程的透明性至
545 关重要。申办者应与监管机构就研究方案的科学合理性及
546 研究实施中有关方案变更的问题进行充分沟通交流。在研
547 究开始前就研究路径的可行性和方案的合理性进行沟通，
548 特别是对数据治理或数据管理计划、样本量确定的依据、

549 以及统计分析计划等进行讨论。在研究的实施过程中，若
550 研究方案有重大调整，即对研究结果有潜在影响的方案修
551 订，例如数据治理计划的变更，统计分析计划中主分析基
552 本模型的变更等，申办方应与监管机构充分沟通，与监管
553 机构达成一致后再进行方案的修订。在研究完成后和递交
554 资料前，申办者应与监管机构就研究结果进行沟通。

555

参考文献

- 556 [1] 国家药品监督管理局. 《真实世界证据支持药物研发与审评的指导原则（试
557 行）》. 2020.
- 558 [2] 国家药品监督管理局. 《用于产生真实世界证据的真实世界数据指导原则
559 （试行）》. 2021.
- 560 [3] 国家药品监督管理局. 《真实世界研究支持儿童药物研发与审评的技术指导
561 原则（试行）》. 2020.
- 562 [4] 国家药品监督管理局. 《基于人用经验的中药复方制剂新药临床研发指导原
563 则（试行）》. 2022.
- 564 [5] 国家药品监督管理局. 《罕见疾病药物临床研究技术指导原则（试行）》.
565 2022.
- 566 [6] 国家药品监督管理局. 《罕见疾病药物临床研究统计学指导原则（试行）》.
567 2022.
- 568 [7] 国家药品监督管理局. 《单臂试验支持上市的抗肿瘤药上市许可申请前临床
569 方面沟通交流技术指导原则（试行）》. 2020.
- 570 [8] 国家药品监督管理局. 《单臂试验支持上市的抗肿瘤药进入关键试验前临床
571 方面沟通交流技术指导原则（试行）》. 2020.
- 572 [9] 国家药品监督管理局. 《药物临床试验亚组分析指导原则（试行）》. 2020.
- 573 [10] Berger M, Daniel G, Frank K, et al. A frame work for regulatory use of real-world
574 evidence[J]. White paper prepared by the Duke Margolis Center for Health Policy,
575 2017.06.
- 576 [11] Boslaugh S . Encyclopedia of epidemiology. [M] SAGE Publications. 2008.
- 577 [12] Chen EY, Raghunathan V, Prasad V. An overview of cancer drugs approved by the
578 us food and drug administration based on the surrogate end point of response
579 rate[J]. JAMA Intern Med .2019;179(7):915-921
- 580 [13] Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials[J]. N Engl J
581 Med, 2017, 377(14): 1391-1398.
- 582 [14] Hernán MA, Robins JM. Causal Inference[M]. Boca Raton: Chapman & Hall/CRC
583 2019.
- 584 [15] ICH E10: Choice of Control Group and Related Issues in Clinical Trials. 2000
- 585 [16] James S. Importance of post-approval real-word evidence[J]. European Heart
586 Journal-Cardiovascular Pharmacotherapy, 2018; 4(1):10-11.
- 587 [17] Last JM. A Dictionary of Epidemiology[M]. 4th Edit. Oxford University Press.
588 2001.
- 589 [18] Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic
590 data[M]. Springer Science & Business Media, 2011.

- 591 [19]Lash TL, Fox MP, MacLehose RF, et al, Good practices for quantitative bias
592 analysis[J], International Journal of Epidemiology, 2014, 43(6): 1969–1985.
- 593 [20]Roland M, Torgerson DJ. Understanding controlled trials: What are pragmatic
594 trials?[J]. BMJ, 1998, 316(7127): 285.
- 595 [21]Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and
596 what can it tell us[J]. N Engl J Med, 2016, 375(23): 2293-2297.
- 597 [22]Strayhorn JM. Virtual controls as an alternative to randomized controlled trials for
598 assessing efficacy of interventions [J]. BMC Medical Research Methodology 2021;
599 21(3): 1-14.
- 600 [23]Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical
601 trials[J]. JAMA, 2014, 311(23)ua ua : 2381-2382.
- 602 [24]US FDA. Framework for FDA’s real-world evidence program. December 2018.
603 2019.
- 604 [25]Velentgas P, Dreyer NA, Nourjah P, et al. Developing a Protocol for Observational
605 Comparative Effectiveness Research: A User’s Guide. AHRQ Publication No.
606 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality;
607 2013. [www.effectivehealthcare.ahrq.gov/ Methods-OCER.cfm](http://www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm).
- 608 [26] Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of
609 Observational Studies in Epidemiology (STROBE)statement: guidelines for
610 reporting observational studies[J]. Annals of internal medicine, 2007, 147(8): 573-
611 577.

612 **附录 1 名词解释**

613 1.初治者（New User）：纳入研究队列之前在充分的洗
614 脱期内未使用研究用治疗的病例。

615 2.单臂临床研究（Single-arm/One-arm Study）：一种只
616 设置试验组或治疗组的非随机临床研究，通常采用外部对
617 照，如历史对照、平行对照或目标值对照等。

618 3.定量偏倚分析（Quantitative Bias Analysis, QBA）：可
619 用于评估研究结果对各种可能的系统误差来源（如错误分
620 类、不受控制的混杂和选择偏倚等）敏感性的一类方法，
621 可基于设置的各类偏倚大小及其分布考察各类偏倚对效应
622 估计结果的影响方向和程度，亦可用于提供进一步校正各
623 种可能偏倚后的分析结果。

624 4.估计目标（Estimand）：对治疗效应的精确描述，反
625 映了针对临床试验目的提出的临床问题。它在群体水平上
626 汇总比较相同患者在不同治疗条件下的结局。

627 5.观察性研究（Observational Study）：又称非干预性研
628 究，根据特定研究问题，不施加主动干预的、以目标人群
629 为对象的、探索治疗与结局因果关系的研究。

630 6.回顾性观察性研究（Retrospective Observational
631 Study）：在研究开始时确定目标人群、并根据历史数据
632 （研究开始前生成的数据）开展的观察性研究。

633 7.模仿目标临床试验（Emulating Target Trial）：是一种

634 真实世界研究方法，即基于现有的真实世界数据，模仿一个
635 个良好的 RCT 设计，产生一个真实世界研究的子集，通过
636 因果推断得出研究结论。

637 8. 偏倚 (Bias)：任何在研究设计、数据收集、分析、
638 结果解释、报告中系统性地导致估计量偏离真值的倾向。

639 9. 前瞻性观察性研究 (Prospective Observational Study)：
640 在研究开始时确定目标人群、并在研究开始前确定将要收
641 集的治疗和结局数据的观察性研究。

642 10. 实用临床试验 (Pragmatic Clinical Trial/Pragmatic
643 Trial, PCT)：又称实操/实效临床试验，指尽可能接近临
644 床真实世界环境的临床试验，是介于RCT和观察性研究之间
645 的一种研究类型。

646 11. 数据治理 (Data Curation)：指针对特定临床研究
647 问题，为适用于统计分析而对原始数据所进行的治理，其
648 内容至少包括数据采集（可包含多个数据源）、数据安全
649 性处理、数据清洗（逻辑判断及异常数据处理、数据完整
650 性处理等）、数据导入和结构化（通用数据模型、归一化、
651 自然语言处理、医学编码、衍生点位等）、数据传输等若
652 干环节。

653 12. 外部对照 (External Control)：在临床试验中，以
654 研究对象以外的数据为对照，以评价所研究的治疗或干预

655 效果。外部对照可以是历史数据，也可以是平行观测所获
656 得的数据，还可以是目标值。

657 13. 虚拟对照方法（Method of Virtual Control）：基于
658 反事实论证思路，首先根据单臂设计所考虑的重要变量，
659 利用现有真实世界数据建立不用试验药物的预后预测模型；
660 然后将单臂试验所获得的协变量数据代入预测模型，计算
661 得到假设不用试验药物的预后预测结果作为对照（虚拟对
662 照）；最后将试验药物实际观测到的结果与虚拟对照进行
663 比较，从而做出是否有效的推断。

664 14. 因果推断（Causal Inference）：基于真实世界数据，
665 刻画干预或治疗与临床结局或健康结局的因果关系路径，
666 充分考虑各种协变量和已测或未测混杂因素的影响，并控
667 制可能的偏倚，采用恰当的统计模型和分析方法，做出干
668 预或治疗与临床结局或健康结局的因果关系的推断结论。

669 15. 真实世界数据（Real-World Data, RWD）：来源
670 于日常所收集的各种与患者健康状况和/或诊疗及保健有关
671 的数据。并非所有的真实世界数据经分析后都能成为真实
672 世界证据，只有满足适用性的真实世界数据才有可能产生
673 真实世界证据。

674 16. 真实世界研究（Real-World Research/Study，
675 RWR/RWS）：指针对预设的临床问题，在真实世界环境下
676 收集与研究对象健康状况和/或诊疗及保健有关的数据（真

677 实世界数据)或基于这些数据衍生的汇总数据,通过分析,
678 获得药物的使用情况及潜在获益-风险的临床证据(真实世
679 界证据)的研究过程。

680 17. 真实世界证据(Real-World Evidence, RWE):指
681 通过对适用的真实世界数据进行恰当和充分的分析所获得
682 的关于药物的使用情况和潜在获益-风险的临床证据。

683 18. 中间变量(Intermediate Variable):指处于因果关
684 系链中间、既受药物治疗影响、同时又影响结局的变量,
685 或与结局有关联的变量;前者又称中介变量(mediator)。

686

687 附录 2 中英文词汇对照

688 中英文词汇对照表

中文	英文
标准操作程序	Standard Operation Procedure, SOP
病例对照研究	Case-control Study
初治者	New User
定量偏倚分析	Quantitative Bias Analysis, QBA
队列研究	Cohort Study
发表偏倚	Publication Bias
工具变量	Instrumental Variable
估计目标	Estimand
观察性研究	Observational Study
恒定时间偏倚	Immortal-time bias
横断面研究	Cross-sectional Study
患者报告结局	Patient Reported Outcome, PRO
回顾性研究	Retrospective Study
回忆偏倚	Recall Bias
混杂因素	Confounder
可追溯性	Traceability
每个协变量阳性事件数	Events per Variable, EPV
碰撞变量	Collider Variable
领先时间偏倚	Lead-time Bias

中文	英文
模仿目标临床试验	Target Trial Emulation
目标临床试验	Target Trial
起点时间偏倚	Zero-time Shift Bias
前瞻性研究	Prospective Study
倾向评分	Propensity Scores, PS
时依变量	Time-varying Variable
实用临床试验	Pragmatic Clinical Trial, PCT
数据管理	Data Management
数据治理	Data Curation
统计分析计划	Statistical Analysis Plan, SAP
幸存者偏倚	Survivor Bias
虚拟对照	Virtual Control
衍生变量	Derived Variable
因果推断	Causal Inference
真实世界数据	Real World Data, RWD
真实世界研究	Real World Research/Study, RWR/RWS
真实世界证据	Real World Evidence, RWE
中间变量	Intermediate variable